

(19) RÉPUBLIQUE FRANÇAISE  
INSTITUT NATIONAL  
DE LA PROPRIÉTÉ INDUSTRIELLE  
PARIS

(11) N° de publication :  
(à n'utiliser que pour les  
commandes de reproduction)

2 782 877

(21) N° d'enregistrement national :

98 10888

(51) Int Cl<sup>7</sup> : H 04 N 7/15, H 04 N 5/232, G 08 C 17/00, 23/00,  
G 06 T 7/00, G 10 L 15/08

(12)

DEMANDE DE BREVET D'INVENTION

A1

(22) Date de dépôt : 31.08.98.

(30) Priorité :

(71) Demandeur(s) : FRANCE TELECOM Société anonyme — FR.

(43) Date de mise à la disposition du public de la demande : 03.03.00 Bulletin 00/09.

(72) Inventeur(s) : VIALLET JEAN EMMANUEL,  
PERAUD RAPHAEL, COLLOBERT MICHEL et BER-  
NIER OLIVIER.

(56) Liste des documents cités dans le rapport de recherche préliminaire : Se reporter à la fin du présent fascicule

(73) Titulaire(s) :

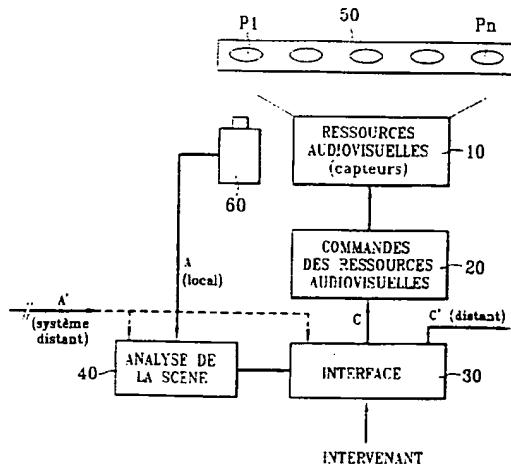
(60) Références à d'autres documents nationaux apparentés :

(74) Mandataire(s) : CABINET BALLOT SCHMIT.

(54) SYSTÈME AUTOMATIQUE DE PRISE DE SON ET D'IMAGES.

(57) L'invention concerne un système automatique de prise de son et d'images notamment pour visioconférence, comportant des moyens de commande (20) de capteurs de prise de vues et de son (10) et des moyens d'analyse de scène (40) pilotant ces moyens de commande (20) pour obtenir un cadrage automatique de la scène filmée. Selon l'invention il est prévu une interface intelligente pour réaliser la sélection (30) d'une personne ou d'un groupe parmi des personnes de la scène filmée, sur commande d'un intervenant, et le cadrage automatique (30) à partir des informations fournies par les moyens d'analyse de la scène (40), sur la personne sélectionnée ou le groupe.

Application à la visioconférence.



FR 2 782 877 - A1



**SYSTEME AUTOMATIQUE DE PRISE DE SON ET D'IMAGES**

L'invention concerne un système automatique de prise de son et d'images notamment pour la visioconférence.

Le besoin de tels systèmes se fait sentir surtout  
5 dans le domaine de la visioconférence, mais peut également être intéressant dans d'autres situations mettant en oeuvre de la prise de vue avec plusieurs personnes par exemple.

Il s'agit donc de systèmes avec lesquels les prises  
10 de son et de vue ne sont pas pilotées par une ou plusieurs personnes (cameraman et ingénieurs du son dans le domaine de la télévision) mais par un équipement automatique.

Dans le cas de la visioconférence, un intervenant a  
15 cependant la possibilité avec certains systèmes équipés d'une télécommande de piloter la prise de vue d'une caméra, mais ces systèmes présentent des inconvénients comme on va le voir dans la suite.

En effet, des systèmes de visioconférence sont  
20 équipés de moyens de prise de vues et de son, disposant d'équipements (caméras et microphones) qui ne sont pas orientables ou dont l'orientation est commandée par l'intermédiaire d'une télécommande.

Dans des systèmes de visioconférence existants,  
25 diverses caméras télécommandables et contrôlables à distance via une communication série sont utilisés.

La télécommande permet de balayer continûment le site et l'azimut de la caméra ainsi que de faire varier continûment le zoom de la caméra. L'orientation de la caméra dans la direction occupée par une personne ou un groupe de personnes est possible, mais malaisée.

Des directions de l'espace (aux nombres de six pour les deux caméras) peuvent être mémorisées par la caméra. La caméra peut être dirigée dans l'une de ces directions par activation d'une touche de la télécommande ou par contrôle du port série. L'intérêt de cette fonction est d'accéder directement à une direction de l'espace sans avoir à agir par combinaison de touches successives (site, azimut). Ainsi, lorsque les positions de l'espace préenregistrées correspondent à des places autour d'une table, l'utilisateur de la télécommande peut simplement passer d'une personne à une autre.

Il existe également des systèmes reposant sur la localisation acoustique selon lesquels l'orientation d'une caméra est pilotée à partir de l'analyse acoustique de la scène.

L'analyse acoustique de la scène est obtenue à partir de plusieurs microphones qui permettent de déterminer la direction des sources sonores, voire des sources de parole. La direction des sources de parole étant identifiée, elles pourraient être sélectionnées une à une, puis être suivies dynamiquement. La fonction Lime Light de Picture Tel, entreprise qui réalise et commercialise des systèmes de visioconférence, est basée sur la localisation acoustique et permet la détection et le suivi d'une source sonore et l'orientation dynamique d'une caméra.

Il existe également des systèmes reposant sur la localisation visuelle de personnes. Cette localisation visuelle est basée sur l'analyse par traitement d'image issue de la caméra..

Si le champ de la caméra est suffisamment large, et le système performant, toutes les personnes présentes seront détectées. A la différence de la localisation

acoustique, la localisation visuelle bénéficie de la permanence des visages à l'image et de l'absence de réflexion (absence de miroirs dans les salles de visioconférence et dans les lieux habituels de prise de vues). Plusieurs techniques peuvent être utilisées séparément ou non pour détecter des personnes : détection de mouvement, détection de la teinte de la peau, détection de formes ovales, détection globale de visages (par réseau de neurones par exemple), détection d'attributs de visages tels les yeux, la bouche.

Un certain nombre de systèmes de détection de visages existe et certains associent une fonction de prise de vue automatique. On peut citer à titre d'exemples les systèmes qui sont décrits dans les publications suivantes :

P1 - M. Collobert, R. Feraud, G. Le Tourneur, O. Bernier, J.E. Viallet, Y. Mahieux, D. Collobert, "LISTEN : a System for Locating and Tracking Individual Speakers", 2nd Int. Conf. On Automatic Face and Gesture Recognition, Killington, USA, October 1996, pp 283-288.

P2 - Hunke, Locating and Tracking human faces with neural network, Technical Report CMU-CS-94-155, 1994.

P3 - Yang, Wu, Waibel, Focus of attention in video conferencing, Technical Report CMU-CS-96-150, 1996.

Cependant, parmi tous les systèmes qui viennent d'être décrits, un grand nombre présente des inconvénients ou de toute façon ne sont pas satisfaisants pour les besoins qui se font sentir actuellement et qui sont précisé dans la suite.

- En effet, les systèmes ayant une fonction de balayage de site, azimut et focale au moyen d'une télécommande présentent les inconvénients suivants :

Pour obtenir un cadre adapté, l'utilisateur doit appuyer successivement sur plusieurs touches de la

télécommande (site, azimut et zoom) ayant d'obtenir le résultat désiré. Cette fonctionnalité est peu pratique, lente, sujette à des erreurs de manipulation et est donc peu utilisée. Elle est de fait impraticable si 5 l'on souhaite adapter périodiquement ou continûment le cadre de la caméra à une personne qui se déplace ou un ensemble de participants qui naturellement bougent. De la même manière, cette fonctionnalité est également impraticable quand on souhaite changer régulièrement de 10 personne.

- La fonction accès par la télécommande à des directions préenregistrées de l'espace présente les inconvénients suivants :

Le premier inconvénient est lié au fait que les 15 positions doivent être préenregistrées. Elles ne peuvent donc être rapidement continûment modifiées.

Le second inconvénient est de supposer que les personnes occuperont bien les positions préenregistrées et qu'elles n'en bougeront pas. Dans la pratique et 20 même en fixant les chaises au sol, on constate que les personnes bougent et donc sont rarement au centre du cadre, voire sortent du cadre si celui-ci est serré sur la personne. Cet inconvénient est manifeste dans le cadre du système de visioconférence où spontanément les 25 personnes sortent du cadre défini par les directions préenregistrées de l'espace.

L'inconvénient suivant est que cette fonctionnalité d'orientation de la caméra dans une direction de l'espace n'associe pas toujours de focale spécifique. 30 Ainsi, l'utilisateur de la télécommande doit actionner les touches de contrôle de la télécommande pour adapter la taille du cadre à la position effectivement occupée par une personne. En effet, si la personne a tendance à sortir du cadre, en passant à une focale plus courte,

cette personne revient dans le cadre sans avoir besoin de changer la direction de l'espace dans laquelle pointe la caméra. Mais alors le cadre n'est plus serré sur la personne et, il y a éventuellement plusieurs personnes dans le cadre.

Lorsqu'il est possible d'associer et de mémoriser une focale à chacune des directions mémorisées de l'espace, cela implique que d'une part, l'utilisateur détermine au préalable la focale qu'il estime adaptée et d'autre part, la mémorise. D'autre part, cette focale enregistrée est fixe et donc non adaptative et, comme pour la direction, plus ou moins simplement modifiable.

La fonctionnalité d'accès à des directions prédéterminées de l'espace peut convenir à certaines situations stables (télésurveillance), mais ne permet pas de s'adapter à une situation particulière. La caméra pointe dans la direction de l'espace, mais ne sait rien du contenu de l'espace, qu'il soit occupé par une personne ou vide.

Un autre inconvénient, secondaire, est le nombre limité à 6 directions de l'espace qui peuvent être mémorisées par la caméra et donc accessibles par la télécommande. Cet inconvénient est en général résolu en mémorisant ces directions dans un ordinateur et en utilisant une télécommande avec un plus grand nombre de touches.

La localisation acoustique présente les inconvénients suivants :

En raison de la présence de sources sonores parasites (bruits de chaises, de portes, d'appareils, ...) superposées à la parole des personnes présentes et en raison de réflexions multiples des

sources acoustiques sur les murs, la localisation acoustique est souvent peu précise.

D'autre part, l'activité acoustique de parole est par nature intermittente (lorsqu'une personne s'arrête 5 de parler pour écouter)..

Enfin, la localisation acoustique est sensible à l'amplitude de la source sonore.

La localisation visuelle présente les inconvénients suivants :

10 Le principal inconvénient de la localisation visuelle est lié à la complexité des algorithmes, à leur rapidité et à leur robustesse. Néanmoins, plusieurs systèmes sont opérationnels soit sur station de travail soit ordinateur personnel (PC) comme les 15 systèmes développés par le déposant, ou comme dans les publications citées précédemment par le déposant.

La fonction de prise de vue automatique d'un groupe de personnes réalisée par le déposant est, à l'usage, particulièrement utile bien que complexe. Le cadre 20 s'adapte en permanence au nombre et la position des participants d'une visioconférence.

Cependant, il apparaît à l'usage, que bien que plus performant et plus satisfaisant que tous les autres, il 25 soit insuffisant car, comme cela a été dit, des besoins se font sentir sur les possibilités d'effectuer à la demande un cadrage automatique sur l'un ou l'autre des participants et de pouvoir revenir tout simplement au cadre général.

L'invention propose pour cela une interface 30 intelligente apte à réaliser la sélection d'une personne (ou d'un groupe de personnes) parmi les personnes de la scène filmée, sur la commande d'un intervenant, et le cadrage automatique à partir des

informations fournies par l'analyse de la scène, sur la personne sélectionnée (ou le groupe de personne).

L'invention a donc pour objet un système automatique de prise de son et d'images notamment pour visioconférence, comportant des moyens de commande de capteurs de prise de vues et de son et des moyens d'analyse de scène pilotant ces moyens de commande pour obtenir un cadrage automatique de la scène filmée, principalement caractérisé en ce qu'il comporte des moyens de sélection d'une personne ou d'un groupe de personnes parmi les personnes de la scène filmée et des moyens de cadrage automatique à partir des informations fournies par les moyens d'analyse de la scène, sur la personne sélectionnée ou le groupe de personne.

Selon une autre caractéristique, les moyens de cadrage comportent une interface logique apte à établir une correspondance entre la personne sélectionnée ou le groupe et les informations de position issues de l'analyse de scène pour fournir aux moyens de commande les informations de position de cette personne ou du groupe par rapport à la scène filmée.

Selon une autre caractéristique, les moyens de sélection comportent une interface physique permettant de sélectionner l'une quelconque des personnes de la scène ou un groupe, pour avoir un cadrage automatique autour de cette personne ou du groupe ou, de sélectionner l'ensemble des personnes pour avoir un cadrage général de la scène.

Selon un mode de réalisation, l'interface physique de sélection est réalisée par une télécommande IR ou électromagnétique.

La télécommande est la télécommande de la caméra d'analyse d'image ou une télécommande d'analyse

universelle, les signaux de commande de ladite télécommande étant reçus et ré-émis par la caméra d'analyse.

5 Selon une variante d'exécution, les signaux de commande de ladite télécommande sont reçus et ré-émis par un émetteur-récepteur au lieu d'être ré-émis par la caméra d'analyse.

10 Selon un autre mode de réalisation, l'interface physique de sélection comprend une interface graphique.

15 L'interface physique comporte en outre un écran qui peut être tactile, sur lequel sont visualisées la scène et les différentes zones sélectionnables.

20 Selon un autre mode de réalisation, l'interface physique de sélection comporte un périphérique d'entrée/sortie d'ordinateur pour sélectionner les zones repérées.

25 L'interface physique de sélection pourra comporter aussi un dispositif de reconnaissance de la parole.

30 Avantageusement, dans le cas d'un système de visioconférence on peut prévoir que les moyens d'analyse de scène reçoivent un signal local pour l'analyse de scène, et que les moyens de sélection permettent de sélectionner une personne ou un groupe de personnes de la scène filmée localement. Les moyens de cadrage automatique utilisent alors les informations issues de la scène filmée localement.

35 Il peut également être prévu, dans le cas d'un système de visioconférence, que les moyens d'analyse de scène reçoivent un signal d'un système distant pour ou correspondant à l'analyse de scène , et que les moyens de sélection permettent de sélectionner alors une personne ou un groupe de personnes de la scène filmée à distance. Les moyens de cadrage automatique permettent dans ce cas, de piloter le

cadrage de la scène filmée à distance, les signaux de commande étant pour cela transportés vers le système distant par exemple avec le signal audiovisuel.

5           D'autres avantages et particularités de l'invention apparaîtront à la lecture de la description qui est donnée à titre indicatif et nullement limitatif et en regard des dessins annexés sur lesquels :

10           - la figure 1, représente un schéma de principe de l'invention,

              - la figure 2, représente un schéma plus détaillé de l'invention,

              - la figure 3, représente un mode de réalisation particulier,

15           - la figure 4, représente un autre mode de réalisation,

              - la figure 5, représente une variante d'exécution du mode de réalisation de la figure 4.

20           Sur la figure 1 on a représenté de façon schématique un système automatique de prise de son et d'images dans lequel on dispose de ressources audiovisuelles 10 pour filmer et capter le son d'une scène 50.

25           Dans le cas d'un système de visioconférence la scène est composée d'une ou plusieurs personnes dites intervenants P1-Pn sur un site, désirant communiquer avec d'autres personnes d'un site distant.

              Les ressources audiovisuelles 10 sont constituées par des capteurs audio et visuels.

30           Les capteurs audio sont par exemple une série de microphones placés à proximité des intervenants. Les capteurs vidéo sont constitués par une ou plusieurs caméras filmant la scène.

Les ressources audiovisuelles 10 sont pilotées par un dispositif de commande 20 classique, apte à fournir les signaux de commande aux capteurs 10 selon les informations reçues en entrée par l'interface 30 comme cela est détaillé dans la suite.

Conformément à l'invention, les informations reçues en entrée sont fournies par l'interface 30 à partir du dispositif d'analyse de scène 40 et de la sélection effectuée par un intervenant.

Le dispositif d'analyse de scène 40 peut être soit audio, soit visuel, soit audiovisuel associé à des capteurs visuels ou audiovisuels.

Dans l'exemple de réalisation qui est donné, ce dispositif est visuel. A cette fin, on utilise une caméra 60 fixe qui permet de fournir le signal servant à effectuer une analyse de la scène visuelle.

Le dispositif d'analyse de la scène comprend la caméra 60 et des moyens de traitement 40 du signal A fourni par cette caméra. Ces moyens sont réalisés par exemple par un micro-ordinateur ou une station de travail équipée d'un programme spécifique d'analyse de scène existant.

De façon plus précise selon l'exemple de réalisation mis en pratique, les visages des personnes présentes dans le champ visuel, sont détectés par un réseau de neurones, puis ledit programme met en oeuvre un algorithme qui suit les visages détectés. D'autres techniques connues peuvent être utilisées.

Selon un autre mode de réalisation, on peut utiliser un dispositif d'analyse de la scène 40 avec une caméra mobile.

Selon un troisième mode de réalisation, on peut utiliser ou réaliser un dispositif d'analyse de la scène utilisant plusieurs caméras fixes ou mobiles.

Les différents capteurs 10 sont commandés par un dispositif de commande 20 qui reçoit des signaux de commande de l'interface 30 selon la présente invention.

Dans le mode de réalisation qui a été réalisé, 5 illustré par le schéma de la figure 2, il s'agit d'un dispositif de commande 20 d'une caméra motorisée 11 qui assure la prise de vue et d'une antenne acoustique 12 qui assure la prise de son.

On va détailler ci-dessous, la prise de vue et de 10 son pour un ensemble de personnes et pour une seule personne, ce qui correspond à des réalisations effectives.

Les mêmes techniques peuvent être mises en oeuvre pour une prise de vue et de son concernant un groupe de 15 personnes ; le groupe est un sous-ensemble de l'ensemble des personnes.

Dans le mode de réalisation préféré qui a été réalisé, l'analyse de la scène est visuelle, c'est-à-dire que l'on détermine la position des personnes mais 20 on ignore si elles parlent. Dans un autre mode de réalisation impliquant une analyse audiovisuelle (incluant donc le son) les dispositifs de prise de son seront sélectionnés à partir des informations audiovisuelles.

25 Lorsque l'utilisateur demande, à travers l'interface 30, une prise de vue sur l'ensemble de personnes P<sub>1</sub>-P<sub>n</sub>, le dispositif de commande 20 commande la caméra de manière à ce que l'ensemble des personnes, 30 présentes dans le champ d'analyse soient cadrées, en respectant les règles de l'art de la prise de vue dans la mesure où les contraintes de la caméra l'autorisent.

Lorsque l'utilisateur demande, à travers l'interface 30, une prise de vue sur une personne en particulier, le dispositif 20 commande la caméra 11 de

manière à ce que la personne, en respect avec les règles de la prise de vue, soit latéralement centrée, que ses yeux soit approximativement au tiers supérieur de l'image par exemple.

5 Dans un mode de réalisation déjà réalisé, la personne sélectionnée étant encadrée par un ou plusieurs voisins, la prise de vue cherche à isoler cette personne des autres à l'image, dans la mesure où les contraintes liées à la caméra et aux règles de la prise de vue l'autorise.

10 Lorsque l'utilisateur demande, à travers l'interface 30 une prise de son sur l'ensemble des personnes, le dispositif 20 commande la prise de son de manière à capter le champ sonore des différents participants. Ce champ sonore peut être obtenu de différentes manières :

15 1) A partir d'un champ large réalisé par l'antenne acoustique.

2) Par sommation des champs sonores issus de différents dispositifs de prise de son (microphones ou antenne acoustique), dispositifs sélectionnés à partir de l'analyse de la scène.

20 Lorsque l'utilisateur demande, à travers l'interface 30, une prise de son sur une personne, le dispositif 20 commande la prise de son de manière à capter le champ sonore des différentes personnes. Ce champ sonore peut être obtenu de différentes manières :

25 1) A partir du champ étroit et orientable de l'antenne acoustique. C'est le mode de réalisation réalisé décrit précédemment.

30 2) A partir d'un ou plusieurs dispositifs de prise de son, dispositifs sélectionnés à partir des informations visuelles ou des informations audiovisuelles.

L'interface 30 permet à l'utilisateur d'obtenir une prise de vue et de son conformes à sa demande (un plan large sur l'ensemble des personnes, un plan serré sur une personne particulière). L'envoi d'une commande 5 issue de l'interface, déclenche la commande d'orientation des capteurs de prise de son et de vue, en fonction de la scène audiovisuelle, analysée par le dispositif d'analyse de scène.

A cette fin, l'interface comporte une interface 10 logique 31 et une interface physique 32.

L'interface logique 31 est, selon un mode de réalisation, constituée par un programme chargé dans le système de traitement du signal d'analyse de scène. Cette interface 15 31 récupère des informations de position des personnes sur la scène et établit une correspondance entre ces informations de position et les informations de sélection données à travers l'interface physique par l'intervenant.

Cette interface 31 interprète (c'est à dire qu'elle 20 décode) les informations reçues pour fournir des signaux de commande de position interprétables par le dispositif de commande 20 pour effectuer le cadrage désiré autour de la personne sélectionnée ou du groupe.

Plusieurs modes de réalisation de l'interface 25 physique 32 sont donnés dans la suite.

Un premier mode de réalisation consiste en une interface graphique 32A sur un micro-ordinateur ou station de travail P comme représenté sur la figure 3.

Avec une souris 320, l'utilisateur choisit 30 d'obtenir une prise de vue et de son sur l'ensemble des personnes de la scène, en cliquant sur une fenêtre nommée "Ensemble", référencé E... Avec une souris, l'utilisateur choisit d'obtenir une prise de vue et de son sur une personne de la scène, en cliquant sur une

fenêtre portant le numéro de la personne désirée P1-Pn ou du groupe de personnes.

Sur cette interface graphique 32A le libellé en chiffre de personnes peut être remplacé par l'image de 5 la personne 321 obtenue par le système d'analyse. Cette image est obtenue soit à un moment fixé par l'utilisateur, soit elle est automatiquement rafraîchie au cours de la réunion.

Une interface graphique 32A avec l'image des 10 personnes 321 est plus ergonomique pour l'utilisateur, car l'interface affiche les prises de vues que peut sélectionner l'utilisateur.

On peut également prévoir un affichage par incrustation sur l'écran de visioconférence, en format 15 réduit, de l'image des personnes présentes avec des numéros, à la visioconférence et, la sélection alors au moyen d'une télécommande.

Dans un autre mode de réalisation, la souris 320 peut être remplacée par un écran tactile et/ou par un 20 dispositif de reconnaissance de la parole R.

Un autre mode de réalisation réalisé, représenté par la figure 4, consiste à détourner l'usage de l'actuelle télécommande 32B de la caméra d'analyse 60 de la scène visuelle. Le détournement et l'utilisation 25 de cette télécommande a été réalisé pour des raisons de facilité et de rapidité de mise en oeuvre.

La télécommande infrarouge 32B est en communication (commandes CDE) avec la caméra d'analyse 60. Cette caméra d'analyse dispose d'un certain nombre de touches dont notamment des touches correspondant à des mémoires 30 de position et une touche "home" H correspondant à la position de repos de la caméra.

Les mémoires de position ne sont pas utilisées en tant que telles pour pointer des directions de

l'espace, mais on n'utilise que le fait que les touches sont activées.

Les positions des mémoires de position sont initialisées par le dispositif, à la position de repos de la caméra. La caméra d'analyse étant fixe dans un des modes de réalisation, le déclenchement des positions 1 à 6 ou de la touche "home" H n'a pas d'effet sur la position de la caméra d'analyse.

Ainsi, en appuyant par exemple sur la touche "home" H, l'utilisateur déclenche via les dispositifs 60, 40, 30 et 20, une prise de vue et de son sur l'ensemble des personnes présentes dans la scène.

Ainsi, en appuyant sur l'une des touches de 1 à 6 correspondant à la mémoire des positions, l'utilisateur déclenche via les dispositifs 60, 40, 30, et 20, une prise de vue sur la personne correspondante (6 personnes maximum dans cette version)..

D'autres télécommandes peuvent être utilisées qu'elles utilisent ou pas la caméra d'analyse en tant que relais de transmission.

On peut par exemple utiliser une télécommande à programmation universelle pour permettre à l'utilisateur de sélectionner la prise de vue et de son.

Dans une phase préparatoire du matériel en vue de mettre en oeuvre l'application, on aura au préalable enregistré dans la télécommande universelle certains des codes de la caméra d'analyse (mémoire de position, "home", ...). Lors de l'utilisation effective de la télécommande 32B, les signaux infrarouges envoyés par la télécommande sont décodés et analysés par l'interface logique 31.

L'utilisation d'une télécommande universelle, permet d'une part, de rendre indisponibles certaines

commandes de la caméra d'analyse (la télécommande universelle n'apprend pas les codes correspondants), et d'autre part, d'avoir des touches dont le libellé correspond mieux à la demande de l'utilisateur : Ainsi, 5 des touches "+" et "-" permettent à l'utilisateur de zapper d'une personne à une autre, comme il a l'habitude de le faire avec les chaînes de télévision.

On va se reporter maintenant au schéma de la figure 5.

10 Afin de ne pas dépendre de la caméra d'analyse en tant que relais entre la télécommande et l'interface 30 et pour offrir à l'utilisateur une palette de commande plus riche, on peut utiliser un dispositif émetteur récepteur 70 du commerce.

15 Cet émetteur-récepteur 70 capte des signaux infrarouges CDE issus de la télécommande 32B et renvoie des codes vers l'interface logique 31, par exemple à travers un port de communication RS232, connecté à l'interface 30.

20 Les précédents modes de réalisation de l'interface permettent à un utilisateur de sélectionner les prises de vue et de son obtenues avec les capteurs physiquement présents dans la salle où se trouve l'utilisateur. Cela est symbolisé par le signal A 25 (local) servant à l'analyse de la scène sur la figure 1.

30 Dans un autre mode de réalisation, les interfaces 31, 32 précédemment décrites permettent de commander les capteurs de prise de vue et de son physiquement présents dans une salle distante (où ne se trouve pas l'utilisateur), la salle avec laquelle il est en visioconférence par exemple.

Ainsi, l'utilisateur participant à une visioconférence, sélectionne et obtient les prises de

vue et de son désirées. Pour cela, le signal A' (distant) pour l'analyse de scène ou correspondant à l'analyse (signal déjà analysé par un système distant dépourvu du signal vidéo) sera appliqué à une entrée du 5 dispositif d'analyse 40.

Dans ce mode de réalisation, les signaux C' émis par la télécommande infrarouge ou par l'interface graphique (et les signaux nécessaires à la constitution des images de l'interface graphique) sont transportés 10 avec l'image, le son et les autres signaux de la visioconférence.

Dans ce mode de réalisation, l'éventuel conflit de commande des capteurs entre la salle locale et la salle distante doit être géré.

## REVENDICATIONS

1. Système automatique de prise de son et d'images notamment pour visioconférence, comportant des moyens de commande (20) de capteurs de prise de vues et de son (10) et des moyens d'analyse de scène (40) pilotant ces moyens de commande (20) pour obtenir un cadrage automatique de la scène filmée, caractérisé en ce qu'il comporte des moyens de sélection (30) d'une personne ou d'un groupe parmi des personnes de la scène filmée et des moyens de cadrage automatique (30) à partir des informations fournies par les moyens d'analyse de la scène, sur la personne sélectionnée ou le groupe.

2. Système selon la revendication 1, caractérisé en ce que les moyens de cadrage comportent une interface logique (31) apte à établir une correspondance entre la personne sélectionnée et les informations de position issues de l'analyse de scène pour fournir aux moyens de commande (20) les informations de position de cette personne ou du groupe par rapport à la scène filmée.

3. Système selon la revendication 1, caractérisé en ce que les moyens de sélection comportent une interface physique (32) permettant de sélectionner l'une quelconque des personnes de la scène ou un groupe, pour avoir un cadrage automatique autour de cette personne ou du groupe, ou de sélectionner l'ensemble des personnes pour avoir un cadrage général de la scène.

4. Système selon la revendication 3, caractérisé en ce que l'interface physique de sélection est réalisée par une télécommande IR ou électromagnétique (32B).

5. Système selon la revendication 4, caractérisé en ce que la télécommande (32B) est la télécommande de la caméra d'analyse d'image (60), les signaux de commande de ladite télécommande (32) étant reçus et ré-émis par la caméra d'analyse (60).

10 6. Système selon la revendication 4, caractérisé en ce que la télécommande (32B) est une télécommande universelle, les signaux de commande de ladite télécommande (32B) étant reçus et ré-émis par la caméra d'analyse (60).

15 7. Système selon la revendication 4, caractérisé en ce que la télécommande (32B) est une télécommande universelle, les signaux de commande de ladite télécommande (32) étant reçus et ré-émis par un émetteur-récepteur (70).

20 8. Système selon la revendication 3, caractérisé en ce que l'interface physique de sélection comprend une interface graphique (32A).

25 9. Système de visioconférence selon la revendication 4 ou 8, caractérisé en ce que l'interface physique comporte en outre un écran (321) sur lequel sont visualisées la scène et les différentes zones sélectionnables.

30 10. Système selon la revendication 8, caractérisé en ce que l'interface physique de sélection comporte un périphérique (320) d'entrée/sortie d'ordinateur pour sélectionner les zones repérées.

11. Système selon la revendication 3, caractérisé en ce que l'interface physique de sélection comporte un dispositif de reconnaissance de la parole (R).

5

12. Système de visioconférence selon l'une quelconque des revendications précédentes, selon lequel les moyens d'analyse de scène reçoivent un signal d'analyse local (A), caractérisé en ce que les moyens de sélection permettent de sélectionner une personne ou un groupe de personnes de la scène filmée localement et en ce que, les moyens de cadrage automatique utilisent les informations de la scène filmée localement.

15

13. Système de visioconférence selon l'une quelconque des revendications précédentes recevant un signal (A') d'un système distant pour ou correspondant à l'analyse de scène, caractérisé en ce que les moyens de sélection permettent de sélectionner une personne ou un groupe de personnes de la scène filmée à distance et en ce que les moyens de cadrage automatique permettent de piloter le cadrage de la scène filmée à distance, les signaux de commande étant transportés vers le système distant.

20

1/3  
FIG. 1

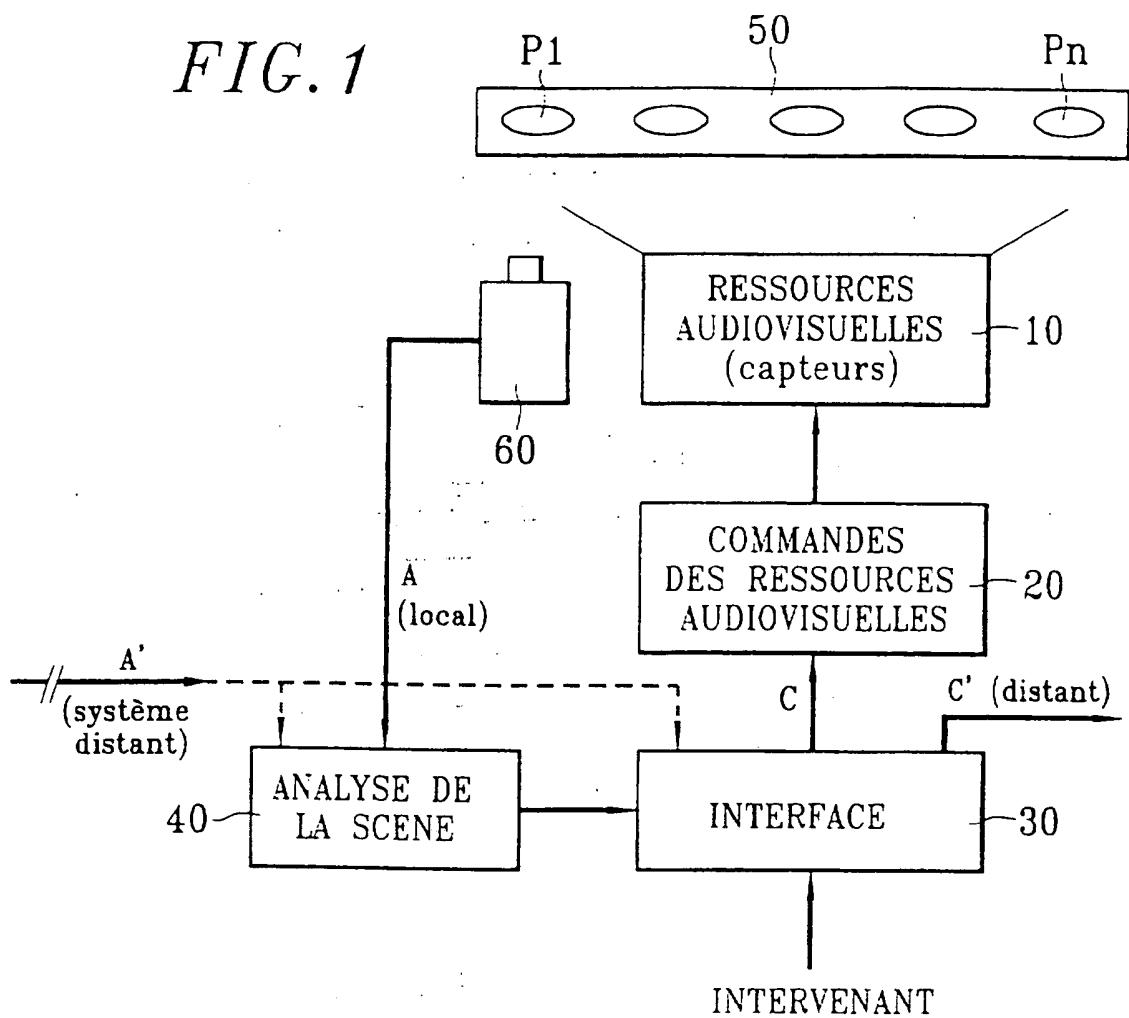
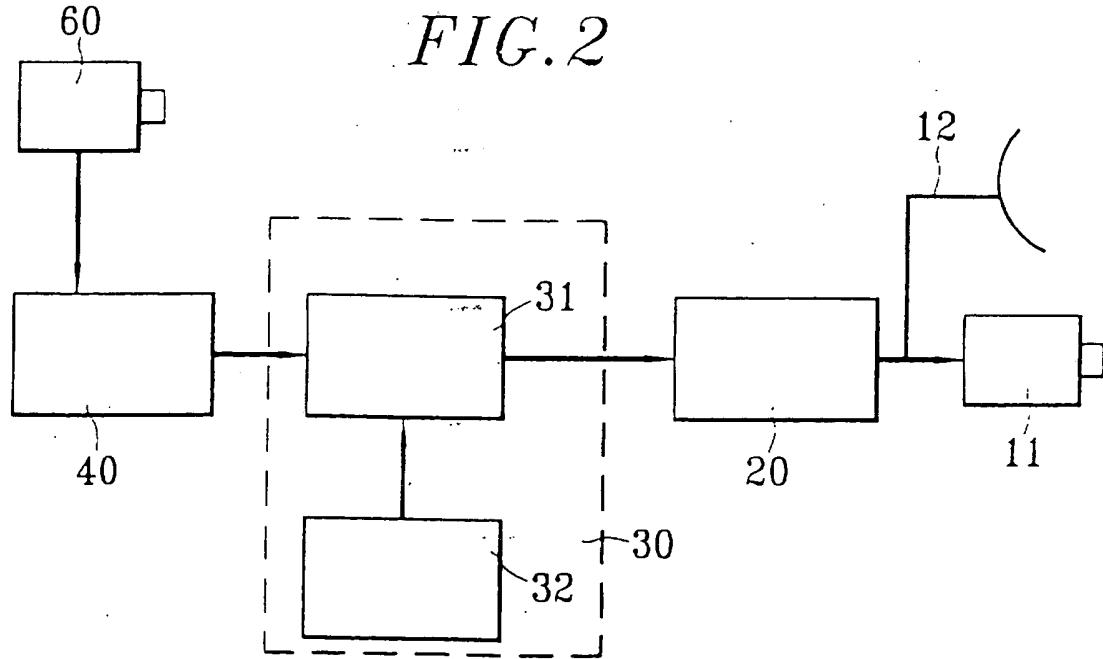


FIG. 2



2/3

FIG.3

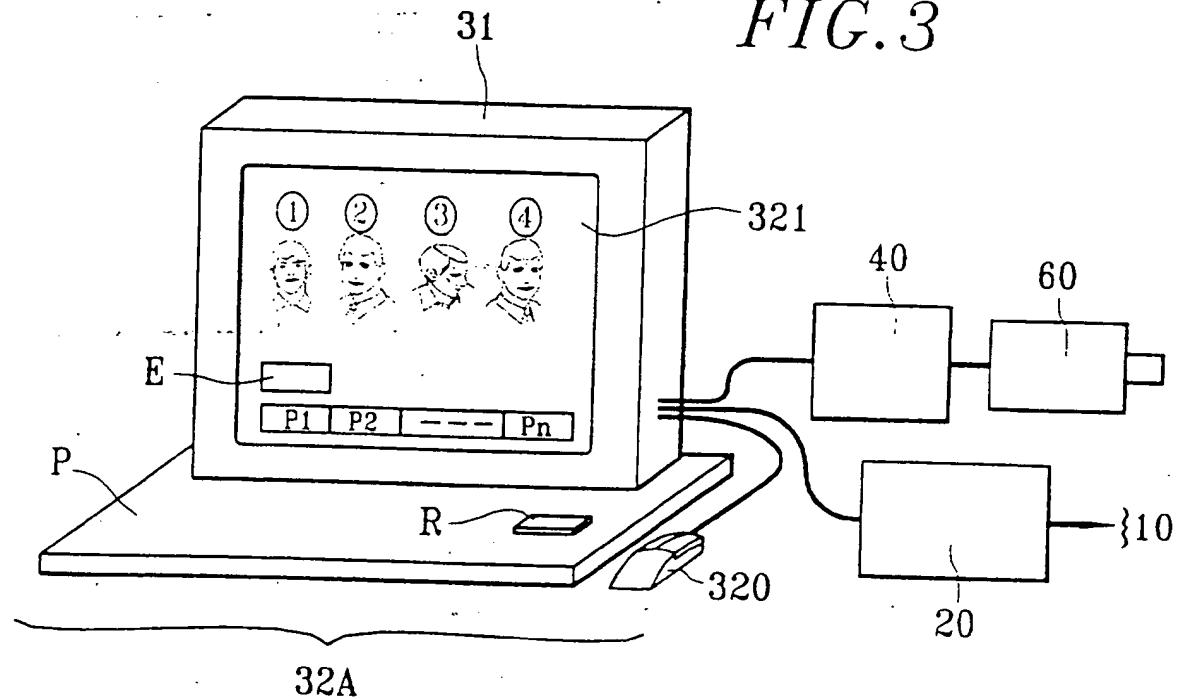
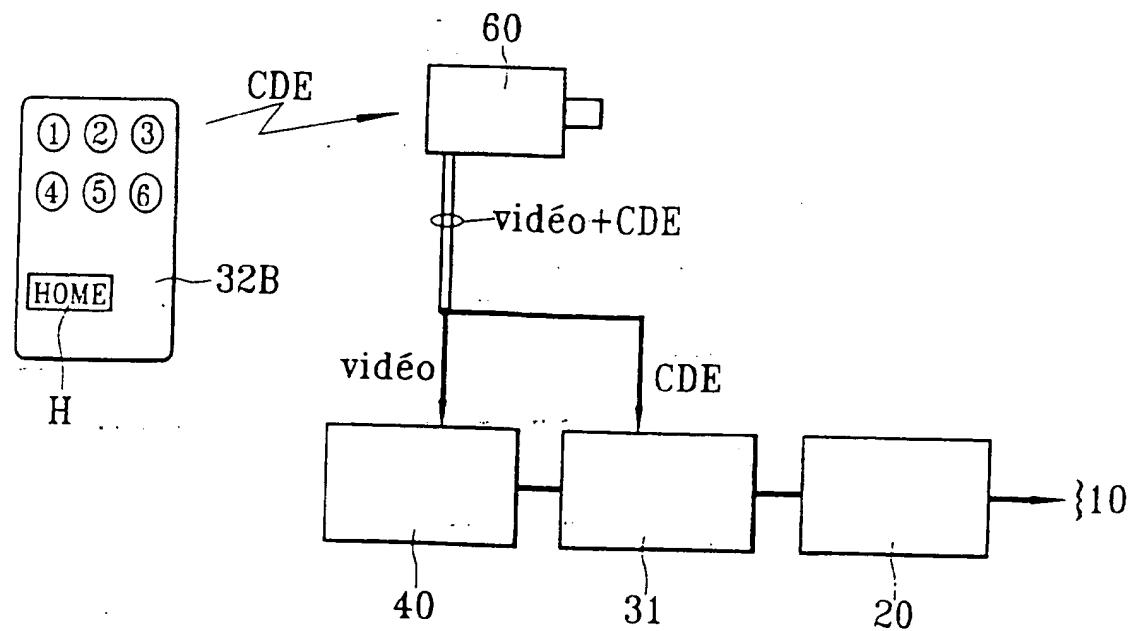
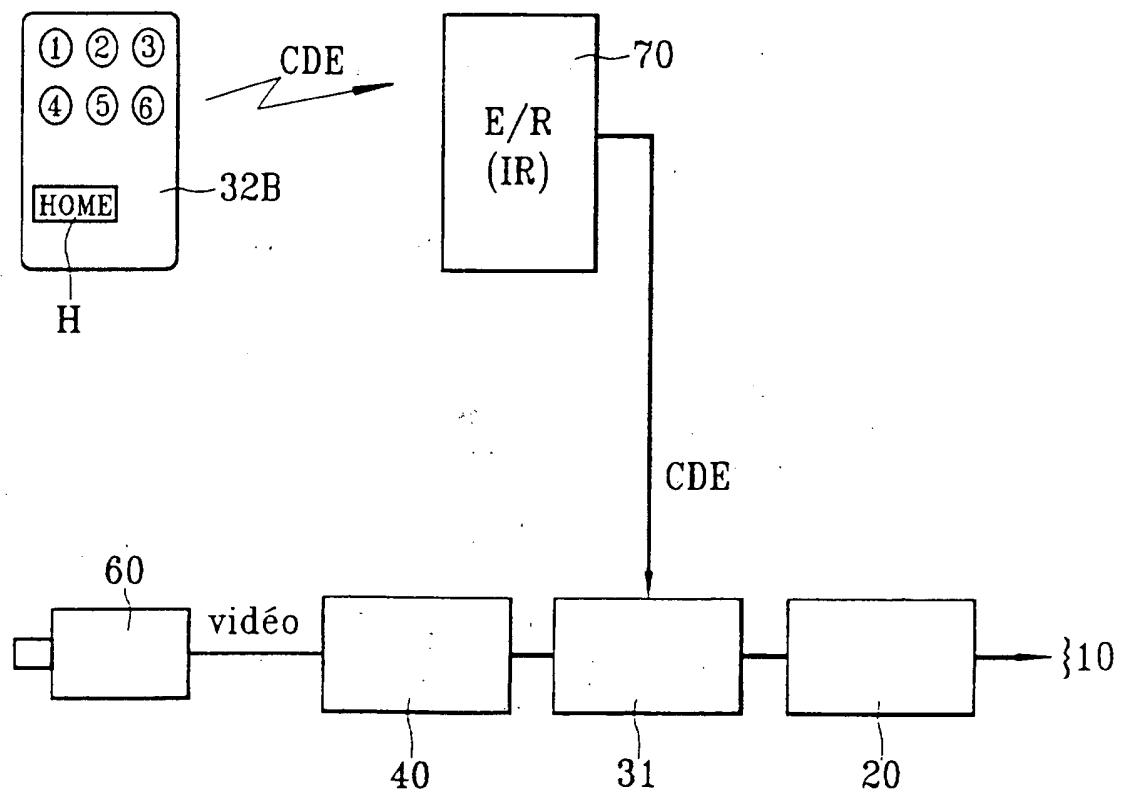


FIG.4



3/3

FIG.5



INSTITUT NATIONAL  
de la  
PROPRIETE INDUSTRIELLE

RAPPORT DE RECHERCHE  
PRELIMINAIRE

N° d'enregistrement  
nationalétabli sur la base des dernières revendications  
déposées avant le commencement de la rechercheFA 563571  
FR 9810888

DOCUMENTS CONSIDERES COMME PERTINENTS		Revendications concernées de la demande examinée
Categorie		
X	WO 96 14587 A (DEVITECH IS :DUE JESPER (DK); HEDER THOMAS (DK); JUERGENSEN PETER) 17 mai 1996	1-3.12
Y	* page 1, ligne 8 - ligne 12 * * page 2, ligne 9 - ligne 14 * * page 3, ligne 16 - ligne 21 * * page 5, ligne 3 - ligne 14 * * page 8, ligne 24 - page 9, ligne 3 * * revendication 7 *	4.8-10
Y	GB 2 252 473 A (RADAMEC EPO LIMITED) 5 août 1992	4.8-10
A	* abrégé * * page 1, ligne 21 - page 2, ligne 18 * * page 4, ligne 12 - ligne 20 *	1-3
A	US 5 434 617 A (BIANCHI MICHAEL H) 18 juillet 1995 * abrégé * * colonne 1, ligne 59 - colonne 2, ligne 6 * * colonne 4, ligne 4 - ligne 8 *	1-3,12, 13
A	WO 95 11566 A (VIDEO CONFERENCING SYS INC) 27 avril 1995 * page 5, ligne 22 - page 6, ligne 24 * * page 15, ligne 32 - page 16, ligne 6 * * page 21, ligne 9 - page 26, ligne 35 *	1-10,12, 13
A	US 5 686 957 A (BAKER ROBERT G) 11 novembre 1997 * abrégé * * colonne 1, ligne 45 - ligne 65 * * colonne 5, ligne 48 - ligne 54 * * colonne 6, ligne 37 - ligne 42 *	1-3,12, 13
		-/-
2	Date d'achèvement de la recherche	Examinateur
	4 mai 1999	Marie-Julie. J-M
CATEGORIE DES DOCUMENTS CITES		
X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : pertinent à l'encontre d'au moins une revendication ou arrière-plan technologique général O : divulgation non-écrite P : document intercalaire		
T : théorie ou principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été divulgué qu'à cette date de dépôt ou qu'à une date postérieure D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant		

INSTITUT NATIONAL  
de la  
PROPRIETE INDUSTRIELLE

RAPPORT DE RECHERCHE  
PRELIMINAIRE

établi sur la base des dernières revendications  
déposées avant le commencement de la recherche

N° d'enregistrement  
national

FA 563571  
FR 9810888

DOCUMENTS CONSIDERES COMME PERTINENTS		Revendications concernées de la demande examinée
Categorie	Citation du document avec indication, en cas de besoin des parties pertinentes	
A	US 5 745 161 A (ITO KAN) 28 avril 1998  * abrégé * * colonne 1, ligne 56 - colonne 2, ligne 24 *	1.3.4. 8-10
A	US 4 286 289 A (OTTESEN ROBERT B ET AL) 25 août 1981  * abrégé * * colonne 1, ligne 62 - ligne 66 * * colonne 2, ligne 27 - ligne 45 *	1.3.4. 8-10
A	US 4 274 609 A (FERRIER NOEL H L ET AL) 23 juin 1981  * abrégé *	1.3.4.8. 10
A	EP 0 751 473 A (LUCENT TECHNOLOGIES INC) 2 janvier 1997  ---	
A	VIALLET J E . COLLOBERT M . FERAUD R . BERNIER O: "Panorama: A what I see is what I want contactless visual interface" LOS ALAMITOS, CA, USA, IEEE COMPUTING SOCIETY, USA, 14 avril 1998. pages 568-572. XP002101885  -----	DOMAINES TECHNIQUES RECHERCHES (Int.CL.6)
2		
Date d'achèvement de la recherche		Examinateur
4 mai 1999		Marie-Julie. J-M
CATEGORIE DES DOCUMENTS CITES		
X	particulièrement pertinent à lui seul	T théorie ou principe à la base de l'invention
Y	particulièrement pertinent en combinaison avec un autre document de la même catégorie	E document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure
A	pertinent à l'encontre d'au moins une revendication ou arrière-plan technologique général	D cité dans la demande
O	divulgation non-écrite	L cité pour d'autres raisons
P	document intercalaire	S membre de la même famille, document correspondant